

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平6-95690

(43) 公開日 平成6年(1994)4月8日

(51) Int.Cl.⁵

G 1 0 L 3/00

識別記号

5 3 1 L

庁内整理番号

7627-5H

E 7627-5H

F I

技術表示箇所

審査請求 未請求 請求項の数1(全4頁)

(21) 出願番号 特願平4-244671

(22) 出願日 平成4年(1992)9月14日

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72) 発明者 松井 知子

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72) 発明者 古井 貞▲熙▼

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

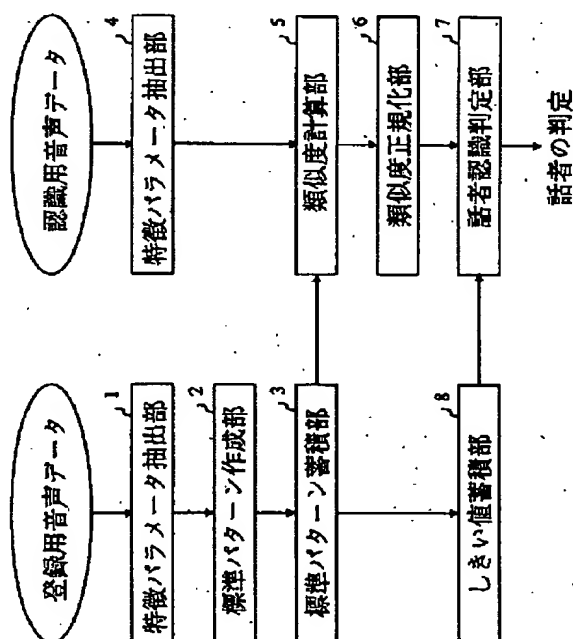
(74) 代理人 弁理士 澤井 敬史

(54) 【発明の名称】 話者認識方法

(57) 【要約】

【目的】 本発明は、申告した話者（以降申告話者という）であると主張する話者の入力音声、特徴パラメータを用いた表現形式に変換し、これと予め話者対応に登録された特徴パラメータを用いた表現形式との類似度を求めて申告話者が本人であるか否かを判定する話者認識方法において、正規化のための特定の言葉を発声することなく、上記類似度の、発声内容、収録時期、伝送系、マイクロホンなどの違いによる変動を吸収して話者認識を行う方法を提供することを目的とする。

【構成】 本発明による話者認識方法は、申告話者の入力音声を、申告話者を含めた複数の話者の特徴パラメータの標準パターンと比較して類似度を計算し、その上位n名（nは1以上の整数）の平均類似度を、申告話者の標準パターンとの類似度から差し引くことによって、上記類似度のばらつきを正規化し、その正規化された類似度を使って本人か否かを判定することを特徴とする。



1

【特許請求の範囲】

【請求項1】 申告した話者（以降申告話者という）であると主張する話者の入力音声、特徴パラメータを用いた表現形式に変換し、その表現形式による入力音声と、予め話者対応に登録された上記表現形式による特徴パラメータの標準パターンとの類似度を求めて、上記入力音声を発声した話者が本人であるか否かを判定する話者認識方法において、

上記入力音声を申告話者を含めた複数の話者の標準パターンと比較して類似度を計算し、その上位n名（nは1以上の整数）の平均類似度を、申告話者の標準パターンとの類似度から差し引くことによって、上記申告話者の標準パターンの類似度の発声内容、収録時期、伝送系、マイクロホンなどによるばらつきを正規化し、その正規化された類似度を使って本人か否かを判定することを特徴とする話者認識方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、例えばインターホンの音声から訪問者は誰であるかを認識したり、入力された音声により暗証番号の人と同一人であることを同定したりするためなどに用いられ、入力音声を、特徴パラメータを用いた表現形式に変換し、その表現形式による入力音声と、予め話者対応に登録された上記表現形式による標準パターンとの類似度を求めて、入力音声を発声した話者を認識する話者認識方法、特にその類似度の正規化方法に関する。

【0002】

【従来の技術】 話者認識は、話者が発声した文章などの音声に含まれる特徴パラメータ（例えばケプストラム、ピッチなど）を求め、登録話者の特徴パラメータの標準パターンとの類似度によって判定する手法がよく用いられる。この類似度は、発声内容、収録時期、伝送系、マイクロホンなどの違いによって大きく変動するために、話者認識性能を低下させてきた。

【0003】

【発明が解決しようとする課題】 従来、この類似度の正規化のために、特定の言葉を話者に発声させ、その音声を申告話者以外の話者の標準パターンに与えて、その音声と申告話者以外の話者の標準パターンとの類似度を計算し、その類似度を使って、本人が認識のために発声した音声と標準パターンとの類似度の値を正規化する方法が試みられてきた（A. Higgins, L. Bahler, and J. Porter, "Speaker verification using randomized phrase prompting", Digital Signal Processing 1, pp. 89-106 (1991)）。しかし、この方法では、特定の言葉を必ず発声する必要があった。

【0004】 本発明は、正規化のための特定の言葉を発声することなく、上記類似度の、発声内容、収録時期、伝送系、マイクロホンなどの違いによる変動を吸収する

2

方法を提供することを目的とする。

【0005】

【課題を解決するための手段】 この発明によれば、入力音声を申告話者を含めた複数の話者の標準パターンに与えて、その音声と申告話者を含めた話者の標準パターンとの類似度を計算し、その上位n名の平均類似度を、申告話者の標準パターンとの類似度から差し引くことによって、上記類似度のばらつきを正規化する。

【0006】

【作用】 本発明による方法は、入力音声と登録話者の標準パターンとの類似度を、正規化のために特定の言葉の発声を要すること無く、正規化することができる。

【0007】

【実施例】 本発明による類似度の正規化の方法を図3を用いて説明する。図3において矩形で囲んだア、イ、ウ、エ、は申告話者を含む複数の話者（ア、イ、ウ、エ、）の標準パターンであり、円形で囲んだA、B、Cは時期A、B、Cにおける申告話者の入力音声である（当然標準パターンおよび入力音声は、特徴パラメータによる表現形式に変換されたものであるが、図3での説明では省略する）。この図は、標準パターンが各話者ごとに異なること、さらに入力音声の標準パターンに対する類似度が時期により異なる（これは発声内容による違いも含む）ことを示している。従って、たとえその入力音声が発声どおり本人のものであったとしても、その本人の標準パターンとの類似度の値は、時期によりばらつき、その類似度の値にしきい値を設定して本人か否かを判定する話者認識の性能を低下させる。

【0008】 本発明による類似度の正規化は以下のように行う。まず入力音声を各話者の標準パターンと比較して類似度を計算し（図の矩形と円形をつなぐ線分が類似度に相当）、その上位n名の平均類似度を求め、入力音声と申告話者の標準パターンとの類似度と、平均類似度との差を求める。平均類似度の値は発声内容、伝送系、マイクロホンなどによる違いをあらわす尺度となっているため、この差は入力音声が発声本人のものである場合には安定して大きく、他の場合には小さくなる可能性が高い。そのために時期がA、B、Cと変化しても安定して本人か否かを判定することができる。

【0009】 次にこの発明の実施例を説明する。この発明では、図1に示すように、登録用音声データの特徴パラメータ抽出部1に入力する。特許パラメータ抽出部1では、入力された音声を例えばケプストラム、ピッチなどの特徴パラメータを用いた表現形式に変換する。次に、特徴パラメータの時系列に変換された登録用音声データが、標準パターン作成部2に入力され、登録用音声データに含まれる特徴パラメータの標準パターンが、例えばベクトル量子化の符号帳、複数のガウス分布の組合せなどで表現される。符号帳あるいは複数のガウス分布の組合せを作成する方法としては、例えば文献「松井知

3

子、古井貞▲熙▼：“VQ、離散／連続HMMによるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89、1991」に述べられている方法などを用いることができる。次に、その標準パターンを標準パターン蓄積部3に蓄える。

【0010】次に、話者を認識する段階では、認識用音声データを特徴パラメータ抽出部4に入力する。特徴パラメータ抽出部4では、入力された音声の特徴パラメータ抽出部1と同じ表現形式に変換する。特徴パラメータの時系列と、標準パターン蓄積部3に蓄えられた申告話者を含む複数の話者の登録用音声データに含まれる特徴パラメータの標準パターンが、類似度計算部5に入力されて、それぞれの類似の度合いが計算される。この具体的方法としては、例えば文献「松井知子、古井貞▲熙▼：“音源・声道特徴を用いたテキスト独立形話者認識”、電子情報通信学会音声研究会資料、SP90-26、1990」に述べられている方法などを用いることができる。計算された類似度の値は、類似度正規化部6に入力される。類似度正規化部6では、上位n名の平均類似度を、申告話者の標準パターンに対する類似度の値から差し引くことによって、申告話者の標準パターンに対する類似度の値を正規化する。nの値は、予め1以上の整数の適当な値に設定しておく。なお、実験的に3名程度に設定すればよいことがわかっている。その正規化された類似度の値は話者認識判定部7に送られ、話者の判定を行う。話者認識判定部7では、しきい値蓄積部8から、その申告話者の声とみなせる類似度の変動の範囲を示すしきい値を読み出して、上記の類似度の値と比較し、その類似度の値が読み出されたしきい値よりも大きければ本人の音声であると判定し、しきい値よりも小さければ他人の音声であると判定する。

【0011】

【発明の効果】以上述べたように、この発明では、申告話者を含めた複数の話者の標準パターンとの類似度を使って、申告話者の標準パターンに対する類似度の値を正

4

規化しており、発声内容、収録時期、伝送系、マイクロホンなどの違いによる類似度の変動の影響を受け難い話者認識を行うことができる。

【0012】次に実験例を述べる。実験は、男性23名、女性は13名が約5ヵ月に渡る3つの時期（時期A、B、C）に発声した文章データ（1文章長は平均4秒）を対象とする。これらの音声で、従来から使われている特徴量、つまり、ケプストラムの細かい時間毎の時系列に変換する。ケプストラムは標準化周波数12kHz、フレーム長32ms、フレーム周期8ms、LPC分析（Linear Predictive Coding、線形予測分析）次数16で抽出した。学習には、時期Aに発声した10文章を用い、テストでは、時期B、Cに発声した5文章を1文章ずつ用いた。

【0013】各話者の標準パターンは、64個のガウス分布の組合せ（「松井知子、古井貞▲熙▼：“VQ、離散／連続HMMによるテキスト独立形話者認識法の比較検討”、電子情報通信学会音声研究会資料、SP91-89、1991」）で表した。結果は平均照合誤り率で評価した。その結果を図2に示す。図2は時期Aを基準とした話者照合の5文章での平均誤り率を示したものである。これより、この発明方法は類似度の正規化を施さない場合と比較して、平均照合誤り率がほぼ一桁小さくなった。以上より、この発明方法は有効であることが実証された。

【図面の簡単な説明】

【図1】図1はこの発明による方法を示すブロック図である。

【図2】図2はこの発明の実験結果を示す図である。

【図3】図3はこの発明における正規化の方法を説明する図である。

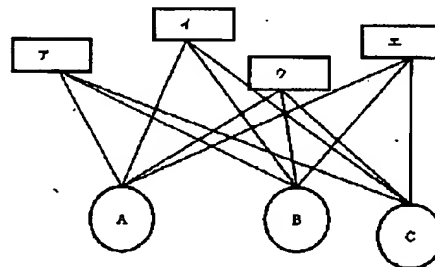
【符号の説明】

- 5 類似度計算部
- 6 類似度正規化部
- 7 話者認識判定部
- 8 しきい値蓄積部

【図2】

正規化	なし	あり					
上位n位	-	1	2	3	5	10	36
時期B	2.8	0.3	0.1	0.2	0.5	0.8	3.1
時期C	3.4	0.3	0.2	0.0	0.2	0.4	2.5
平均	3.1	0.3	0.2	0.1	0.4	0.6	2.8

【図3】



【図1】

